

Web Results 1 - 10 of about 7,330 for (term or phrase or word) and query and (sum or total) and weight. (1)

CONTAINS Query Operators

... qualifier in a broader or narrower **term query**, the qualifier ... an acceptable substitution for a **word** in a ... documents that contain either the **phrase** alsatians are ...

www.cise.ufl.edu/help/database/oracle-docs/text.920/a96518/cqoper.htm - 101k - [Cached](#) - [Similar pages](#)

[PPT] ISP433/633 Week 3

File Format: Microsoft Powerpoint 97 - [View as HTML](#)

... Context queries. **Phrases**. Proximity. Boolean queries. ... **Search Words**. Optional field or index qualifications. Boolean Operators. ... **Query expansion**. **Term re-weighting**. Type. ...

www.albany.edu/faculty/hy973732/isp433/notes/ISP433w3.ppt - [Similar pages](#)

[PPT] Mandarin-English Information (MEI): Investigating Translingual ...

File Format: Microsoft Powerpoint 97 - [View as HTML](#)

... Bilingual. **Term**. List. **Query**. ... Inquiry #**sum()** operator. TDT-2, **phrase-based translation**, **word-based retrieval**. Retrieval Granularity. Character bigrams are best. ...

www.glue.umd.edu/~oard/papers/queens.ppt - [Similar pages](#)

[PDF] Information retrieval models Query methods

File Format: PDF/Adobe Acrobat - [View as HTML](#)

... return documents that contain the **phrase** "UFO Sightings ... return documents – that contain the **word** k a ... documents containing this **term** The **term weight** is given ...

www.cs.rpi.edu/~sibel/mmdb/lectures/ir_models.pdf - [Similar pages](#)

[PPT] www.ist.temple.edu/~vucetic/cis670_fall2002/irmodesl.ppt

File Format: Microsoft Powerpoint 97 - [View as HTML](#)

... Missing syntactic information (eg **phrase** structure, **word** order, proximity ... Given a two-term **query** "AB", may prefer a document containing A ...

[Similar pages](#)

Understanding Query Expressions

... Thesaurus Operators. The thesaurus operators expand a **query** for a single **term** (**word** or **phrase**) using a thesaurus that defines relationships between the user ...

www-rohan.sdsu.edu/doc/oracle/context206/A54630_01/ch03.htm - 101k - [Cached](#) - [Similar pages](#)

CS397CXZ Assignment #2: Pivoted Normalization vs. BM25 (Okapi) ...

... the average length of documents in the collection, the **total** counts of a **term** in the ... Group a **word** pair that occurs frequently as one single **phrase**. ...

sifaka.cs.uiuc.edu/course/397cxz03f/assign2.html - 12k - [Cached](#) - [Similar pages](#)

[PDF] Queries in Oracle 9i Text

File Format: PDF/Adobe Acrobat - [View as HTML](#)

... can be either single **words** or **phrases** and must ... queries when the expression has more than one **query term**. ... documents that have the more similar **words** compared to ...

nordbotten.ifi.uib.no/VirtualMuseum/Publications/OracleTextQueries-Nina-draft.pdf - [Similar pages](#)

[PDF] Discriminative Power and Retrieval Effectiveness of Phrasal ...

File Format: PDF/Adobe Acrobat - [View as HTML](#)

... Consider a supplemental phrasal **term** as informative if ... Single **words Phrases** Single

words + phrase Short query 2.81 ... rel)/p(occ)) value for query terms {#phrasal ...
terral.lsi.uned.es/irnlp2000/papers/fujita.pdf - [Similar pages](#)

[PDF] [Okapi Chinese text retrieval experiments at TREC Introduction ...](#)

File Format: PDF/Adobe Acrobat - [View as HTML](#)

... the effect of different **phrase** weighting functions ... approach perform and better than
the **word** approach city ... assign the usual **sum** of individual **term** weights to ...

research.microsoft.com/users/robertson/papers/trec_pdfs/okapi_trec6_chinese.pdf - [Similar pages](#)

Gooooooooogle ►

Result Page: 1 2 3 4 5 6 7 8 9 10 [Next](#)

(term or phrase or word) and que

[Search within results](#) | [Language Tools](#) | [Search Tips](#) | [Dissatisfied? Help us improve](#)

[Google Home](#) - [Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

©2004 Google

Information retrieval models

- Documents and queries are characterized by a number of index terms
 - Based on a query (representation of an information problem), guess the relevance of each document
 - Rank documents in the order of relevance
 - Return the most relevant documents
- The effectiveness of an IR system depends on the ability of the document representation to capture the “meaning” of the documents with respect to the users’ needs

Multimedia Database Systems - Spring 2001

1

Query methods

- Browsing
- Adhoc retrieval
 - Document collection remains stable, users try to find relevant documents using adhoc queries
- Filtering
 - User queries remain stable as “profiles”
 - As new documents are added they are sent to users who might be interested in these documents
 - Profiles can be constructed on keyword queries, terms occurring in documents retrieved by users

Multimedia Database Systems - Spring 2001

2

1

Information retrieval model

- An information retrieval model is a quadruple $\langle D, Q, F, R(q_i, d_j) \rangle$ where
 - D is a set composed of logical views (or representations) for the documents in the collection
 - Q is a set composed of logical views (or representations) for the user information needs called “queries”
 - F is a framework for modeling document representations, queries and their relationships
 - $R(q_i, d_j)$ is a ranking function which associates a real number with a query q_i in Q and a document representation d_j in D.

Multimedia Database Systems - Spring 2001

3

Documents

- A document is a collection of words
- An index term is an “important” word that
 - Possess a meaning, such as a noun and has been simplified (stop words, stemming)
 - Distinguishes the document from the others
- The set of all index terms for a document collection is given by $\{k_1, \dots, k_t\}$
- A document d_j in IR is usually given by a vector:
$$d_j = \langle w_{1,j}, \dots, w_{t,j} \rangle$$
 where $w_{i,j}$ is the weight of term k_i in document d_j .

Multimedia Database Systems - Spring 2001

4

2

Documents

- Assumption:
 - The occurrence of a term t_1 in a document is completely independent of the occurrence of another term t_2 in the same document
 - Not true in general, but does not appear to have a big impact on the retrieval effectiveness

Boolean model for retrieval

- A Boolean query contains query terms connected by logical connectives and, or, not.
- A Boolean query is interpreted as a set membership function.
- Query:
 - Q = “UFO” return documents that contain the word “UFO”
 - Q = “UFO Sightings” AND “Albany” return documents that contain the phrase “UFO Sightings” and the word “Albany”

Boolean model for retrieval

- $Q = k_a \text{ and } (k_b \text{ or not } k_c)$ return documents
 - that contain the word k_a and
 - either contain k_b or does not contain k_c
- In the boolean model, each document either
 - satisfies the query, then we return 1 (relevant)
 - does not satisfy the query, then we return 0 (irrelevant)
- Documents can be represented as a vector of 0s and 1s
 - 1 if a term appears and 0 if it does not appear

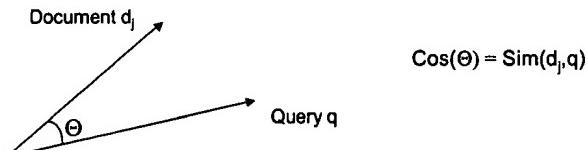
Multimedia Database Systems - Spring 2001

7

Vector model

- In the vector model, both queries and documents are weighted vectors
- The relevance of a document to a query is given by the “cosine of the angle” between a document vector and a query vector

$$\text{Sim}(d_j, q) = \sum_{i=1..l} (w_{ij} \cdot w_{iq}) / \sqrt{\sum_{i=1..l} (w_{ij}^2) \cdot \sum_{i=1..l} (w_{iq}^2)}$$



Multimedia Database Systems - Spring 2001

8

Vector model

- The importance of a term in a document depends on:
 - How important it is for identifying the content of this document (term frequency)
 $f_{i,j} = \text{freq}_{i,j} / (\max_i \text{ freq}_{i,j})$
frequency of term k_i in document d_j , versus
the highest frequency of a term in the
same document
 - How important it is for identifying the document from the others
(document frequency)
 $\text{idf}_i = \log N/n_i$
total number of documents versus
total number of documents containing this
term
- The term weight is given by $f_{i,j} * \text{idf}_i$

Multimedia Database Systems - Spring 2001

9

Vector model

- A user query consists of a number of terms
- How do we assign weights to query terms:

$$w_{i,q} = (.5 + (.5 \text{ freq}_{i,q} / \max_i \text{ freq}_{i,q})) . \log N/n_i$$

Multimedia Database Systems - Spring 2001

10

Fuzzy set model

- A fuzzy set has a membership function, $\mu_A(u)$, that returns a real number $0 \leq \mu_A(u) \leq 1$.
 - If $\mu_A(u) = 1$, then A is definitely a member
 - If $\mu_A(u) = 0$, then A is definitely not a member
- Fuzzy sets use a number of pre-set functions to determine the meaning of various connectives
 - $\mu_{\text{not } A}(u) = 1 - \mu_A(u)$
 - $\mu_{A \text{ or } B}(u) = \max \{\mu_A(u), \mu_B(u)\}$ or $\mu_A(u) + \mu_B(u)$
 - $\mu_{A \text{ and } B}(u) = \min \{\mu_A(u), \mu_B(u)\}$ or $\mu_A(u) * \mu_B(u)$

Multimedia Database Systems - Spring 2001

11

Fuzzy set model

- Determine the term-to-term correlation in a collection of documents between terms k_i and k_l

$$c_{i,l} = n_{i,l} / (n_i + n_l - n_{i,l}) \quad \text{where } n_x \text{ is the number of documents containing term } k_x$$

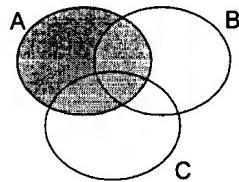
Then, compute $\mu_{i,j} = 1 - (\prod_{k \in d_j} (1 - c_{i,k}))$
the degree of membership of document d_j to term k_i

Multimedia Database Systems - Spring 2001

12

Fuzzy queries

- Given a query $q=k_i$ then similarity of a document d_j to q is given by μ_{ij}
- Given a query $q = k_i \text{ AND } k_j$, the similarity of a document d_j to query q is given by $\mu_{ij} * \mu_{kj}$ (or using any appropriate operator for AND)
- Similarly for OR (use + or max)
- Given a complex query: $(A \text{ and } (\text{not } B)) \text{ or } (C)$,

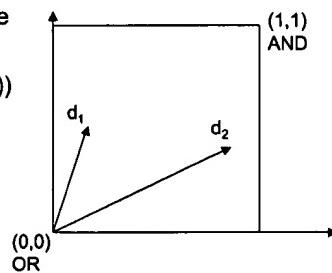


Multimedia Database Systems - Spring 2001

13

Extended Boolean Model

- Suppose, you are given a query containing keywords k_x and k_y
- Assume, the weight of terms k_x and k_y in document d_j are given by (x_1, y_1)
- Given query " $k_x \text{ OR } k_y$ ", we would like to be as far away from $(0,0)$ as possible hence maximize $\text{distance}((0,0), (x_1, y_1))$
- Given query " $k_x \text{ AND } k_y$ ", we would like to be as close to $(1,1)$ as possible hence maximize $1 - \text{distance}((1,1), (x_1, y_1))$



Multimedia Database Systems - Spring 2001

14

Extended Boolean Model

- Under this model:
 - $\text{Sim}(\text{or-query}, d) = \sqrt{(x^2+y^2)/2}$
 - $\text{Sim}(\text{or-query}, d) = 1 - \sqrt{((1-x)^2+(1-y)^2)/2}$
- Suppose now connectives and/or have a degree "p"
 - I.e. or-query: $k_1 \text{OR}^p k_2 \text{OR}^p \dots \text{OR}^p k_m$
 - $\text{sim}(\text{or-query}, d) = \text{power}((x_1^p+x_2^p+\dots+x_m^p)/m), 1/p)$
 - I.e. and-query: $k_1 \text{AND}^p k_2 \text{AND}^p \dots \text{AND}^p k_m$
 - $\text{sim}(\text{and-query}, d) = 1 - \text{power}(((1-x_1)^p+(1-x_2)^p+\dots+(1-x_m)^p)/m), 1/p)$

Multimedia Database Systems - Spring 2001

15

Extended Boolean Model

- Given p-norms, we have the following properties:
 - If $p = 1$, then $\text{sim}(\text{or-query})=\text{sim}(\text{and-query})= (x_1+\dots+x_m)/m$
 - Reduces to arithmetic mean
 - If $p = \infty$, then $\text{sim}(\text{or-query})= \min(x_k)$ and $\text{sim}(\text{and-query}) = \max(x_k)$

Multimedia Database Systems - Spring 2001

16